Análisis probabilístico en evaluación de riesgos

M.Sci. Laura Pruzzo

Introducción

En el marco determinístico que hemos revisado recientemente, el incremento en el riesgo de cáncer de por vida, para un individuo expuesto a un químico y a una vía en particular, es una función de la concentración y otras variables de exposición, y de la potencia carcinogénica; todas estas variables representadas por valores puntuales. Esta formulación ignora la definición fundamental del riesgo, como la *probabilidad* de un resultado adverso.
Puede ocurrir que la distorsión sea aún mayor, si todos los valores que se seleccionan son altamente conservadores. Al combinarlos multiplicativamente, podrían dar lugar a un estimador puntual poco realista. Si el valor resulta inferior al nivel regulatorio, tendremos confianza en que se trata de un riesgo aceptable; pero si resulta superior, ¿se trata de un riesgo inaceptable, o es producto de la distorsión por la utilización de valores extremos? Y si bien se ha demostrado que los estimadores puntuales tradicionales pueden resultar muy conservadores y exagerar el riesgo involucrado, existen ejemplos de lo contrario, en los cuales un nivel de análisis probabilístico puede establecer objetivos más estrictos en términos de limpieza de sitios contaminados.
En conclusión, debemos tener presente que el valor selecto para representar una variable particular, no es más que uno de los tantos posibles valores que pueden observarse .También hemos visto que pueden existir distintos criterios acerca de que valor puntual elegir: de tendencia central, o de máxima exposición.

Finalmente es importante recalcar que el nivel de análisis determinístico o puntual, es importante y no debe omitirse pues es la herramienta para la investigación de antecedentes. Puede resultar suficiente, y no todo sitio peligrosamente contaminado requerirá de un nivel posterior de análisis probabilístico. Si las posibilidades de mitigación son accesibles, tendrá más sentido limpiar el sitio, que realizar un análisis probabilístico.

1. El marco probabilístico

Dado que el riesgo se define como la probabilidad de un evento adverso, la teoría de probabilidades, la estadística y el álgebra de variables aleatorias proveen las herramientas naturales para cuantificar riesgos. Veremos entonces, los conceptos básicos de probabilidad. Hacia el final del curso, veremos que es posible aproximar la distribución resultante de riesgo mediante el método de simulación Montecarlo, pero antes deberemos conocer algo sobre las distribuciones más utilizadas en la evaluación.

Adicionalmente en otra entrega abordaremos el enfoque bayesiano para tratar problemas de riesgos a la salud, así como sus aplicaciones al análisis de decisión.

2. Beneficios y costos del abordaje probabilístico

Beneficios:

1. Su resultado es una distribución del riesgo potencial.
2. Utiliza toda la información disponible sobre variabilidad e incertidumbre.
3. Evita problemas derivados de la utilización de valores muy extremos.
4. Permite caracterizar el dictamen de los expertos consultados. Esto es, el analista debe desarrollar distribuciones a partir de la consulta a expertos.

Costos:

1. Se necesita recopilar mucha más información! Y la información, a veces simplemente no existe. Se deben realizar más mediciones y obtener más datos para describir una variable.
2. Se requiere definir una distribución de probabilidad para dicha variable.
3. Las decisiones de manejo del riesgo pueden dificultarse por una mayor cantidad de información resultante.
4. Es más difícil la comunicación de los resultados, en términos que todos los grupos de interés involucrados puedan entender.

3. Las distribuciones de probabilidad

Una variable aleatoria (*v.a.*) es una variable que puede tomar algún valor entre un rango de valores, con una determinada probabilidad de ocurrencia.
El rango de valores que una variable puede tomar y su probabilidad asociada están codificadas en una función matemática llamada la *función de probabilidad* de la variable. Este rango se captura con dos funciones: la función de probabilidad y la función de distribución acumulada (si la v.a es discreta, la primer función se refiere como *función masa de probabilidad* y si es continua, como *función de densidad*). Ambas representan la misma distribución, pero proporcionan distinta información.

Es de fundamental importancia la identificación de las Distribuciones de Probabilidad apropiadas a los datos de entrada. A menudo, esto requerirá analizar datos empíricos o históricos, y ajustados a una distribución; en otras oportunidades, tales datos no están disponibles y se deberá elegir una distribución apropiada mediante la opinión de expertos. Esta consulta debe utilizarse cuidadosamente, es una inferencia individual basada en la evidencia disponible. La utilización de dictamen de expertos, involucra el proceso de construir una distribución de probabilidad subjetiva.
Existen dos definiciones de la noción de probabilidad: objetiva y subjetiva. La definición frecuentista considera que la probabilidad es el límite de frecuencias relativas (o proporciones) de eventos observables. Pero la probabilidad puede también ser el resultado de una construcción mental del observador, que corresponde al grado de "certeza racional" que tenga acerca de una afirmación; como en este marco las probabilidades pueden variar de una persona a otra, se le llaman probabilidades subjetivas o grados de creencia.

Las distribuciones de las variables de exposición y riesgo deben captar variabilidad e incertidumbre. Por ejemplo en el caso del *peso corporal*, la variable aleatoria captura principalmente la variabilidad en una población, conocida y bien medida. O bien, en el caso de la v.a. "*días que una persona va a la pileta*", puede capturar una conducta individual poco conocida, o directamente desconocida.

En este nivel de análisis las distribuciones se seleccionan sobre la base de las características de la situación. Por ejemplo, podría determinarse que, en el algoritmo de exposición a un contaminante del agua, la variabilidad en concentración estaría bien descripta por una distribución triangular con un mínimo de 80, una moda de 85 y un máximo de 120 microgramos/litro; la variabilidad en tasa de ingestión, estaría bien descripta por una distribución normal con una media de 1,60 y una desviación estándar de 0,20 litros/día; la variabilidad en peso corporal, resulta bien caracterizada por una distribución normal con una media de 70 y una desviación estándar de 10 kg, etc. Y con esta información se realiza el análisis.

4. Distribuciones Discretas

Procesos Estocásticos

Un proceso estocástico es un sistema de eventos contables en el cual todos los eventos ocurren de acuerdo a un mecanismo definido. Existen tres procesos fundamentales: Binomial, Poisson e Hipergeométrico.

Gran cantidad de problemas en Análisis de Riesgo pueden abordarse con un buen conocimiento de estos tres procesos.

Proceso Binomial

Sistema de conteo aleatorio en el cual hay *n* ensayos independientes e idénticos, cada uno de los cuales tiene la misma probabilidad de éxito *p* que produce *s* aciertos. Entonces, tenemos tres cantidades {*n*, *p*, *s*}, que describen completamente el proceso binomial.

La función masa de probabilidad de la Distribución Binomial (*n*, *p*) es:

$$p(x) = \binom{n}{x} p^{x} (1-p)^{n-x}$$

La media es *n* . *p*, y la varianza es *n* . *p* (1 - *p*).

Una aplicación de la Distribución Binomial es la estimación de la probabilidad de al menos un acierto, en un conjunto de ensayos. Si denotamos:

$$q = (1 - p)$$

La probabilidad de no tener ningún acierto es q^n , y de tener al menos uno será:

$$P_1 = 1 - q^n = 1 - (1 - p)^n$$

Mac Diarmid (1991) presentó esta aplicación, utilizada por la Oficina de Recursos Rurales de Australia en la evaluación del riesgo asociado a la importación de carne porcina, proveniente de países infectados con gastroenteritis transmisible (GET). En esta expresión, la probabilidad *T* de introducción de GET se relaciona con la probabilidad *p* de que un corte de carne contenga el virus, y el número de ocasiones *n* en que se consumen cortes:

$$T = 1 - (1-p)^n$$

Encontramos ejemplos de aplicación de la distribución Binomial en problemas de sanidad animal, como por ejemplo, detección de enfermedades en importación y exportación, pruebas de sueros o vacunas, etc. Un concepto básico en esta área, es el de *prevalencia* (*p*), o tasa de infección, que representa la proporción de animales o un grupo o población, infectados con una enfermedad en particular.

Proceso Poisson

Proceso en el que existe una oportunidad continua y constante de ocurrencia de un evento. Por ejemplo, en cualquier pequeño intervalo de tiempo durante una tormenta, hay una probabilidad de que ocurra un rayo. Aquí el continuo es el tiempo, pero en Análisis de Riesgo nos puede interesar otro continuo, la exposición. Así, si en un lago hay una distribución aleatoria de quistes de Giardia, el consumo de quistes por parte de turistas acampantes sería un proceso Poisson, en el cual el consumo de agua es la medida de exposición. Es decir, hemos considerado un proceso de distribución de eventos aleatorios en el tiempo, pero el mismo mecanismo se aplica a la distribución de puntos en el plano o en el espacio, como por ejemplo, la ocurrencia de colonias de bacterias en una caja de Petri.

En un proceso Poisson se asume que no hay límite en la cantidad de ocurrencias, que éstas son independientes, y que el número promedio de ocurrencias es constante. Esta última cantidad, es el parámetro clave λ, número medio de eventos que ocurren por unidad de exposición.

La función masa de probabilidad de una variable aleatoria "*x*" que tiene distribución Poisson es:

$$p(x;\lambda) = \frac{\lambda^x e^{-\lambda}}{x!}$$

Con media y varianza igual a λ.

En muchas aplicaciones, tratamos con ensayos binarios donde *n* es comparativamente grande y *p* es pequeña, mientras que el producto *n*p* es de magnitud moderada. En casos así, es conveniente usar una aproximación de la Distribución Binomial a una Distribución Poisson con $\lambda = n\,.\,p\,\left(n \geq 20;\ p \leq 0,05\right)$. Los supuestos subyacentes a las dos distribuciones son similares, esto es: una serie de *n* eventos binarios que ocurren independientemente con probabilidad *p*. Por ejemplo, si *p* = 0,03 y *n* = 100, la probabilidad de *x* = 2 ocurrencias específicas en 100 eventos es:

$$P\left(x=2\right)=\binom{100}{2}\,p^{2}\,\left(1-p\right)^{98}=0,2252; \text{ D. Binomial y}$$

$$P\left(x=2\right)=\frac{e^{-\lambda}\,\lambda^{2}}{2\,!}=\frac{e^{-3}\,3^{2}}{2}=0,2240 \text{ D. Poisson}$$

Así, cuando se considera un evento poco frecuente, por ejemplo la ocurrencia de algún defecto congénito por exposición a un contaminante en el agua consumida por mujeres embarazadas, se usa la D. Poisson para modelar la ocurrencia de dichos eventos.

Existen muchas otras aplicaciones del proceso Poisson en campos tales como la Física, donde permite describir el comportamiento de partículas radioactivas; en Investigación Operativa, el comportamiento del tráfico de aeropuertos; también, la presencia de un microorganismo en un volumen unitario de cierta sustancia, etc.

Proceso Hipergeométrico

Ocurre cuando se muestrea aleatoriamente sin reemplazo de una población – en cuyo caso los ensayos no son independientes- contando el número de elementos de esa muestra, que presentan una característica en particular.

Consideremos un grupo de M ítems individuales, D de los cuales tienen cierta característica. Si se eligiera aleatoriamente *n* ítems sin reemplazo, con la misma probabilidad de ser elegidos cualquier de los “M” del grupo, se trata de un proceso *hipergeométrico*, en el que la probabilidad de observar *x* elementos con la característica de interés está dada por la siguiente función masa de probabilidad:

$$p\left(x\right)=\frac{\binom{D}{x}\binom{M-D}{n-x}}{\binom{M}{n}} \qquad \begin{array}{r} 0 \leq x \leq n \\ x \leq D \\ n \leq M \end{array}$$

A diferencia del proceso binomial, la probabilidad de éxito cambia en cada ensayo dependiendo del resultado anterior, a consecuencia del muestreo sin reemplazo.

La distribución hipergeométrica se aplica a las *encuestas de captura-recaptura*, utilizadas en estudios ambientales para estimar el número real (M, desconocido) de animales en un sitio. Supóngase que se capturan 20 tigres (D), se identifican con caravanas y se liberan. A continuación se realiza una segunda captura de 30 tigres (n), observándose siete (x) individuos con caravana. En la práctica se observan los valores de D, n y x, pero M es desconocido. Sólo sabemos con certeza que será mayor o igual a (20 + 30 - 7) = 43, y podríamos suponer que existen tigres caravaneados en la muestra en la misma proporción que existen en la jungla (x/n proporcional a D/M). Sobre esta base se estima una distribución de “creencia” y no una verdadera distribución de probabilidad, dado que existe un número exacto de tigres.

En muchas situaciones la población es muy grande en comparación a la muestra y podemos asumir que, si la muestra volviera a la población, hay una probabilidad muy pequeña de que vuelva a elegirse. En tal caso, cada muestra tendría la misma probabilidad de elegir un individuo con una característica en particular, deviniendo en un proceso binomial. Esta aproximación, no puede efectuarse si la población es menos de diez veces el tamaño de la muestra. Es decir, la condición es $n < 0{,}1$ M para utilizar una aproximación Binomial (n, D/M).

5. Distribuciones continuas más utilizadas en análisis de riesgo

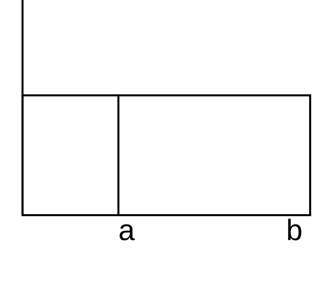
Las distribuciones continuas se definen por su *función de densidad*; en tanto que las discretas, por su *función masa de probabilidad*. Ambas dependen de parámetros que controlan la forma, ubicación, etc. de la distribución.

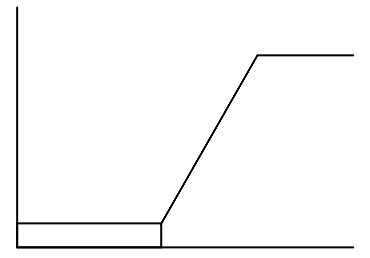
Uniforme

Caracteriza una variable aleatoria para la cual *cualquier resultado entre un mínimo y un máximo es igualmente probable.*

Provee una de las formas más simples de representar nuestra incertidumbre en un modelo. Su uso es apropiado cuando se puede identificar un rango de valores posibles, pero no se puede decidir qué valores de este rango son más probables de ocurrir que otros.

$$f(x) = \frac{1}{b-a} \;\; \forall\, a \le x \le b$$





$$f(x)=\begin{cases} 0 \text{ si } x<a \\ \dfrac{x-a}{b-a} \quad a\le x\le b \\ 1 \text{ si } b<x \end{cases}$$

0

a b

$$u=\frac{a+b}{2}; \qquad \sigma^2=\frac{(b-a)^2}{12}$$

La cantidad "*a*" es un parámetro que controla la locación sobre el eje *x*. La cantidad (*b* – *a*) determina la escala de la distribución; si aumenta, la forma de la función de densidad se alarga y viceversa si disminuye.

Ejemplos de aplicación de la distribución uniforme

- Los números aleatorios generados por computadoras provienen de una distribución uniforme con *a* = 0 y *b* = 1. Esta aplicación es particularmente relevante en el método de simulación Montecarlo.
- Cuando existe poco conocimiento de la variable, los parámetros *a* y *b* se eligen reflejando la mejor estimación del experto acerca del rango de dicha variable.

Normal

Se trata de una distribución no limitada, si bien hemos visto que la mayor proporción de la densidad está cerca de la media. Sus parámetros son μ (locación) y σ^2 (dispersión). Para muchas variables (longitudes, pesos, concentraciones) es teóricamente inapropiada porque permite valores negativos.

Una aplicación importante de esta distribución resulta como consecuencia del Teorema del Límite Central; se cumple que, la media de un grupo de variables aleatorias de cualquier distribución tiene Distribución Normal. La función de densidad Normal se detalla en el Anexo.

Triangular

Para ciertas variables, se asume que los valores hacia el medio del rango son de ocurrencia más probable que los valores cercanos a los extremos. En estos casos, la distribución triangular provee una forma conveniente de representar la incertidumbre. A menudo, es utilizada como "aproximación gruesa" de otra distribución, en ausencia de datos más completos. Es flexible para modelar varios supuestos, pero al ser limitada elimina la posibilidad de valores extremos que podrían ocurrir.

Tiene los siguientes parámetros:

a = mínimo (locación en eje *x*)

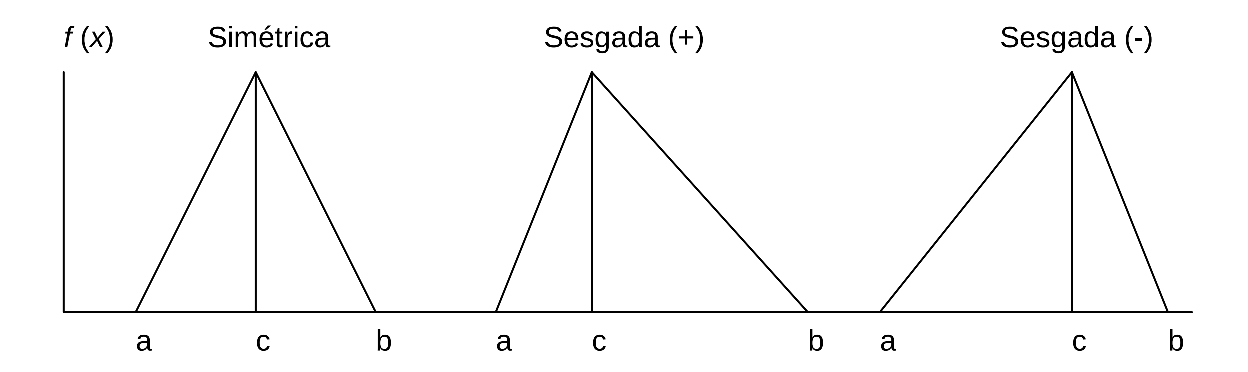
b = máximo (b – a = escala)

c = valor más probable (forma de la distribución)

$$\mu = \frac{a+b+c}{3}\ ;\ \sigma = \frac{a^2+b^2+c^2-ab-ac-bc}{18}$$

En el apéndice se detalla la forma de la función de densidad y de la función de distribución.

Gráficamente se observa que la distribución triangular puede ser:



Exponencial

Utilizada para modelar eventos que se repiten al azar en el tiempo, por ejemplo, tiempo que transcurre hasta la ocurrencia de falla de un mecanismo. Modela un sistema “sin memoria”: la ocurrencia actual no tiene efectos en resultados posteriores. Está limitada por el valor cero donde tiene la mayor densidad, y ésta declina a mayores valores de *x*. Es decir, la moda de la distribución está cercana al cero y la probabilidad de ocurrencia continuamente decrece a mayores valores de la variable.

$$f(x) = \lambda\, e^{-\lambda x} \quad \text{"} \quad x \geq 0$$

$$F(x) = 1 - e^{-\lambda x} \quad x \geq 0$$

$$\mu = 1/\lambda \ ;\ \sigma = (1/\lambda)^2$$

Ejemplo: tiempo que transcurre entre dos tormentas fuertes.

Lognormal

Una variable se distribuye Lognormal cuando su logaritmo tiene distribución normal. Esto es, si X se distribuye Lognormal, entonces Y=ln X se distribuye Normal.

Esta distribución es útil para modelar variables que ocurren naturalmente y que son el producto de otras variables, debido a la aproximación a la normalidad de la suma de los logaritmos. La base mecanística de multiplicación de un gran número de variables aleatorias se observa por ejemplo, en la v.a. “concentración de un tóxico en sangre”, que puede depender del producto de: la cantidad de alimento consumido, la concentración del tóxico en el alimento, la fracción absorbida y el tiempo de permanencia en el cuerpo.

- Como las ecuaciones básicas de Riesgo son multiplicativas, la distribución del riesgo suele ser lognormal

En la práctica, ajusta muy bien datos de concentración de tóxicos en distintos medios. En general, debido a su forma (sesgada (+) y limitada por el valor cero) encuentra aplicación en fenómenos que:

- No pueden tener valores negativos.
- Presentan valores altos con baja probabilidad

Asimismo, muchas distribuciones de otros factores de exposición son bien descriptas por la distribución Lognormal. Entre ellos podemos citar la tasa de respiración por kg de peso, el consumo de agua, la duración de la ducha, el consumo de los principales alimentos, el uso de productos con químicos en minutos por año, etc.

En el apéndice se detalla la forma de la función de densidad.

Estudio de caso 1

La distribución de la concentración de radón en el interior de las viviendas

La figura 1 muestra una distribución empírica de la concentración de radón en los hogares, basada en datos de más de 1.000 hogares en 38 estados de los Estados Unidos. Se observa que:

- Los niveles de radón no pueden ser negativos, por lo tanto aún si los datos tuvieran forma de distribución normal, ésta sería truncada.
- La presencia de radón es generada por un número de eventos interrelacionados, no una suma de eventos independientes.
- La distribución empírica es claramente no simétrica; en cambio, los datos se agrupan en un área y tienen una extensa cola hacia valores altos de concentración.

Por lo tanto, la distribución lognormal puede proveer un buen ajuste ya que:

- Presenta valores > 0.

- Describe elementos “raros” de un conjunto a través de la cola larga de baja probabilidad.

En 1990, otros autores sugirieron que una distribución exponencial ajusta bien los mismos datos (figura 2). ¿Se preferiría una u otra?

Una cola más extensa de la distribución exponencial sugiere que puede haber un número mucho mayor de hogares de muy alto nivel de radón, que el predicho por la distribución lognormal. Si esto fuera así implicaría un uso distinto de los recursos en la identificación y mitigación en aquellos hogares con mayor riesgo.

Figura 1 (Nero et al, 1987)

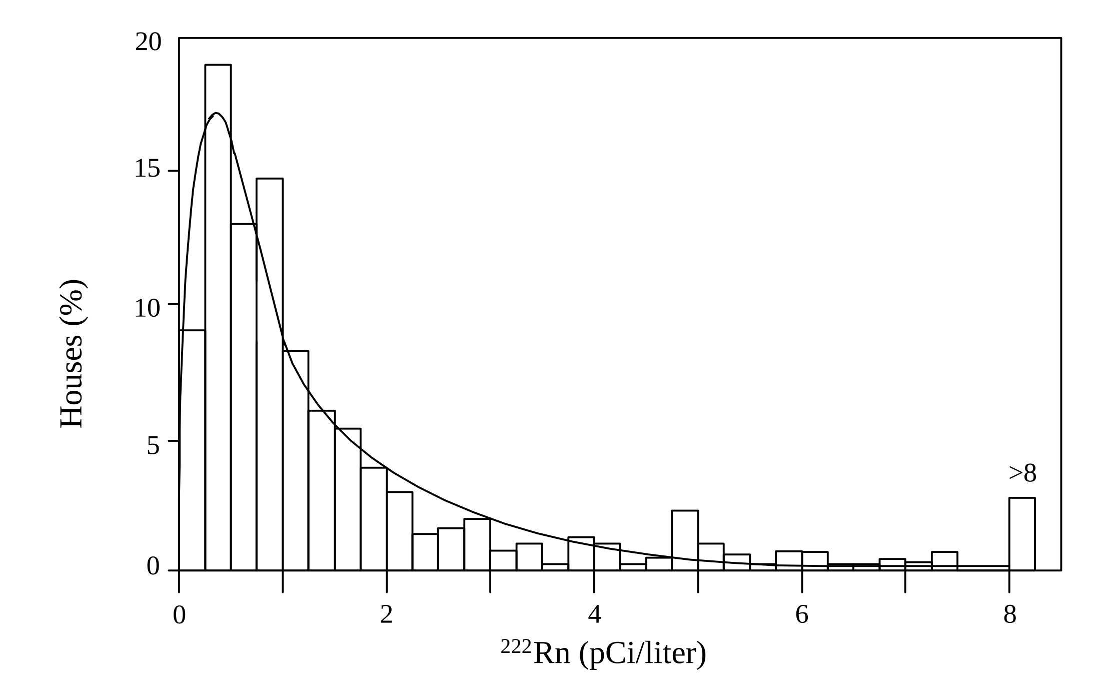
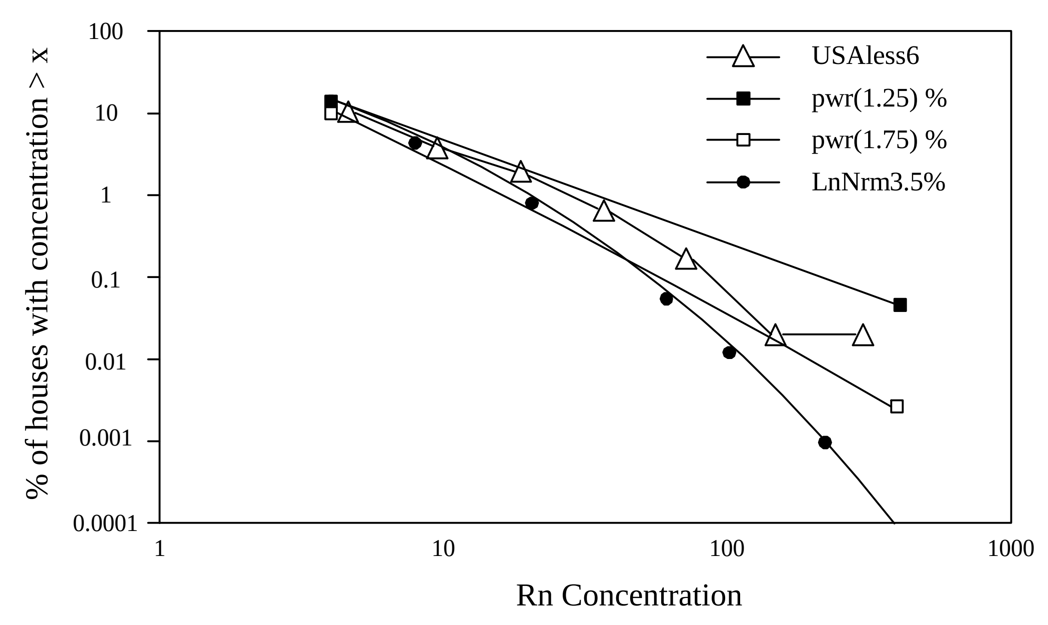


Figura 2 (Goble y Sokolow, 1990)



Beta

La familia de distribuciones beta es una familia continua en el intervalo [0, 1], es decir para $0 < x < 1$; indexada por dos parámetros de forma, α y β, ambos positivos:

$$f(x) = \frac{(\alpha + \beta - 1)!}{(\alpha - 1)!\ (\beta - 1)!}\ x^{\alpha - 1}\ (1 - x)^{\beta - 1}$$

Los valores de α y β se estiman de una muestra con media *m* y varianza s^2

Si $\alpha = \beta \Rightarrow$ simétrica

Si $\alpha < \beta \Rightarrow$ sesgada +

Si $\alpha > \beta \Rightarrow$ sesgada (-)

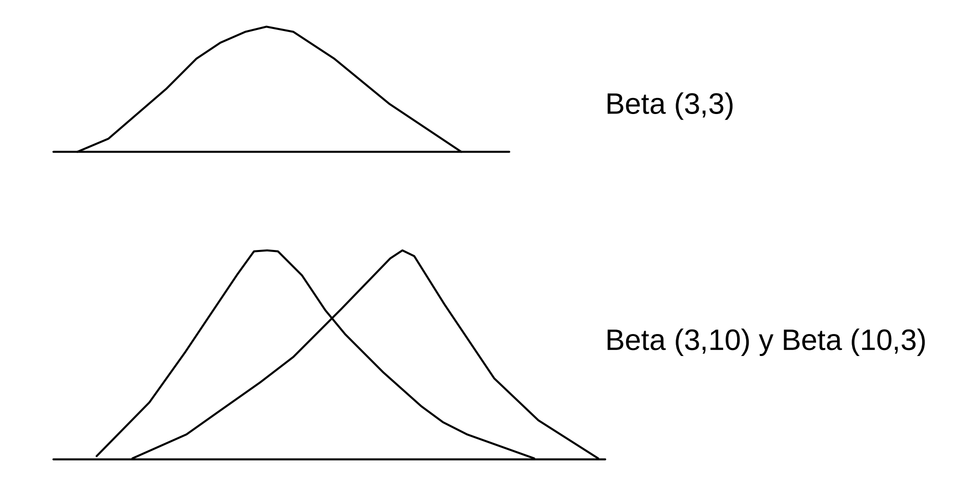
Si $\alpha = \beta = 1 \Rightarrow \mathbf{U}\ (0,1)$
En este último caso, la distribución Beta coincide con la uniforme en (0,1).

Puede corroborarse que media y la varianza de la Distribución Beta son, respectivamente:

$$\mu = \frac{\alpha}{\alpha + \beta} \Rightarrow \text{si } \alpha = \beta \Rightarrow \mu = \frac{1}{2}$$

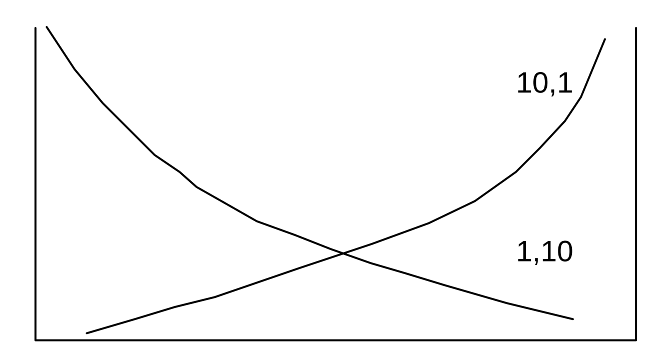
$$\sigma^2 = \frac{\alpha\,\beta}{(\alpha + \beta)^2\ (\alpha + \beta + 1)}$$

Las figuras siguientes permiten apreciar parcialmente la diversidad de formas que puede asumir la distribución Beta, según los valores que adopten α y β.



Beta (3,3)

Beta (3,10) y Beta (10,3)



En el análisis cuantitativo de riesgo la distribución Beta ha encontrado importantes aplicaciones; principalmente en la inferencia bayesiana, donde existe la necesidad de tener una densidad lo suficientemente flexible para modelar proporciones que naturalmente ocurren entre 0 y 1, como el parámetro de la distribución binomial, que puede ubicarse en cualquier punto del intervalo (0,1). En esta situación las probabilidades a priori no se pueden enumerar, pues son infinitas. El modo de representarlas es a través de una función de densidad correspondiente a una distribución continua. Cualquier función de densidad continua definida en el intervalo (0,1) puede ser útil (siempre que refleje el punto de vista del investigador), pero la más usual en el caso en que se estima una proporción es la distribución beta. Dicha distribución se extiende solo sobre el intervalo (0,1), de modo que cumple con la condición óptima para representar nuestra percepción acerca de cuán probables son unos u otros valores de un parámetro que también está constreñido a tales límites.

Estudio de caso 2

La distribución Beta - un modelo consistente para la exposición a contaminantes del aire

Michael R. Flynn. Stoch. Env. Res. and Risk Ass. (2004)18: 306-308

Introducción

El uso de funciones de densidad de probabilidad (PDF) para modelar la concentración de sustancias tóxicas en el aire tiene una larga historia en la evaluación de la exposición ocupacional. La caracterización de mediciones de concentración mediante distribuciones de probabilidad, es la base para efectuar inferencias estadísticas y es el corazón de la evaluación de riesgo y de los estudios epidemiológicos.
Actualmente, la distribución Lognormal parece ser la preferida en exposición ocupacional, debido a su capacidad para ajustar los datos experimentales y sus convenientes propiedades matemáticas. Sin embargo, esta distribución no es

un modelo teórico consistente de los datos de concentración, dado que siempre da una probabilidad finita para valores que son físicamente imposibles. El impacto de este error es difícil de evaluar, y a menudo se asume como insignificante. Sin embargo un modelo alternativo, la distribución Beta, es consistente y suficientemente flexible para acomodar mediciones de concentración.
La noción de consistencia en la representación matemática de un proceso implica que el modelo se ajusta a los límites del proceso considerado. Para representar concentraciones, una pdf debe limitarse a valores entre 0 y 1.
La aplicación de la distribución Beta a contaminantes del aire no parece ser muy común, si bien se citan varios antecedentes.

Teoría

La concentración de un contaminante del aire, es la masa del mismo contenida en un dado volumen de la mezcla de aire y contaminante. Para gases y vapores, es tradicional usar el volumen de contaminante dividido por el volumen de mezcla y multiplicado por un millón, para obtener las partes por millón (ppm). Es claro de esta definición, que por tratarse de una fracción, la concentración está limitada entre 0 y 1, y por encima del máximo la probabilidad es nula.

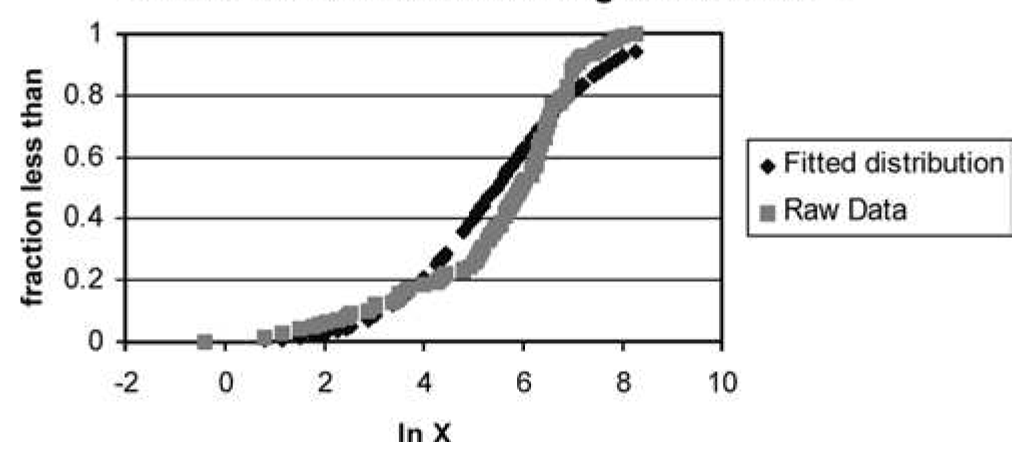
Las distribuciones de contaminantes aéreos son a menudo sesgadas. La mayoría de las observaciones tienden a tener valores relativamente pequeños, en tanto que los valores altos, poco frecuentes, producen la típica cola hacia la derecha.

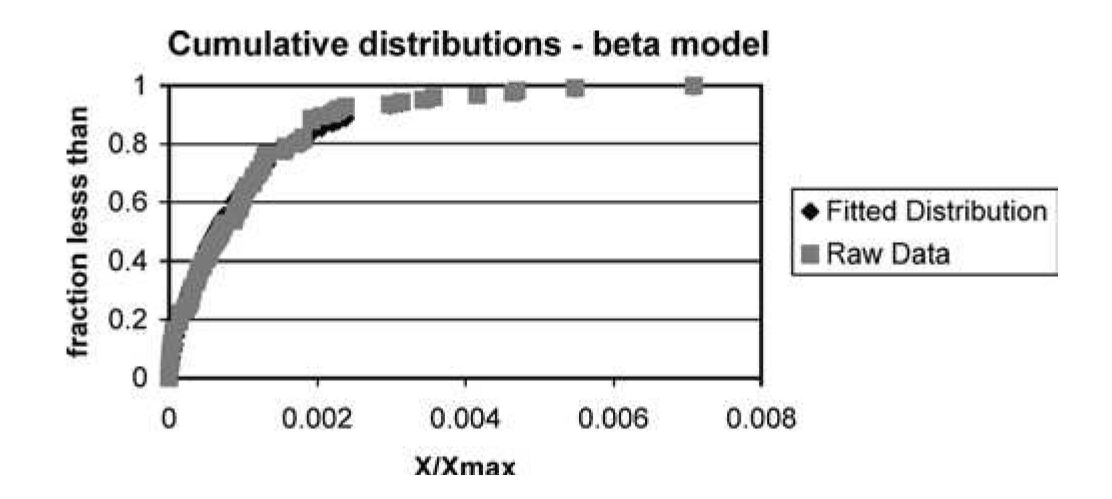
Métodos y resultados

Se analizó un archivo de datos de 126 mediciones de naftaleno en µg/m3. La figura 1 muestra el ajuste con una distribución Lognormal, y la figura 2 el ajuste con una distribución Beta con parámetros $\alpha = 0,76$ y $\beta = 744,3$. Un estadístico X^2 estandar se calculó para los ajustes lognormal ($X^2= 2,15$) y beta ($X^2= 0,66$) indicando el ajuste superior de la distribución Beta.

Conclusión

La distribución de probabilidad Beta es un descriptor matemático flexible de los datos de exposición. Es una distribución continua en el intervalo [0,1] lo cual es consistente con la definición y límites de los datos de concentración. La distribución puede incluir información acerca de la máxima concentración posible o probable, y en algunos casos puede proveer una alternativa razonable al modelo Lognormal. Exposiciones a sustancias con presiones de vapor relativamente bajas como el naftaleno, pueden ser particularmente candidatos para la distribución Beta, dado que el efecto de la cola finita puede ser más significativo.


Cumulative distributions - lognormal model
fraction less than
1
0.8
0.6
0.4
0.2
0
-2
0
2
4
6
8
10
ln X
Fitted distribution
Raw Data


Cumulative distributions - beta model
fraction less than
1
0.8
0.6
0.4
0.2
0
0
0.002
0.004
0.006
0.008
X/Xmax
Fitted Distribution
Raw Data

Trabajo Práctico 4. Distribuciones de Probabilidad

Ejercicios

1. Un experto en higiene y seguridad del trabajo afirma que uno de cada diez accidentes en la ruta se debe a la fatiga de los choferes. ¿Cuál es la probabilidad que al menos tres de cinco accidentes se deban a esta causa?

2. Los oficiales de cuarentena de un país importador saben que hay un 2 % de prevalencia de una enfermedad en un país, del cual un empresario pretende importar 1200 vacunos. Cuántos estarán infectados?

3. Se estima que el 40% de ratones utilizados en un experimento para determinar potencia de un tóxico mueren una hora después de administrarles cierta dosis. Determine la probabilidad de que seis de quince ratones de un laboratorio mueran luego de administrarles la dosis en cuestión.

4. Se asume que la probabilidad de que una persona que asiste a un evento en un día caluroso sufra un golpe de calor, es de 0,05. ¿Cuál es la probabilidad de que 18 personas sufran golpe de calor, en un evento que convocará 3000 personas?

5. En una región desértica, el número de animales de una especie en extinción que enferman por año, por ingestión de una planta tóxica, tiene una D. Poisson con $\lambda = 1,6$. Determine la probabilidad de:
   a) dos animales enfermos en un año.
   b) al menos siete animales enfermos en cinco años.

6. Se capturan 100 gansos salvajes. Los servicios de sanidad animal estiman que en la población total, existe un 3% de prevalencia de la “enfermedad del pico”. ¿Cuántos gansos tendrán la enfermedad?

7. Como parte de un análisis de riesgo por contaminación del aire, se examina la emisión de gases de seis de los 24 camiones de una empresa. Si cuatro de ellos presentan una emisión excesiva de contaminantes, que probabilidad hay de que no resulten incluidos en la muestra de análisis?

8. Entre los 120 aspirantes para ocupar un empleo, sólo 80 están realmente calificados para hacerlo. Si se seleccionan al azar cinco de estos aspirantes para una entrevista, determinar la probabilidad de que sólo dos de los cinco estén calificados, mediante a) la D. Hipergeométrica y b) la D. Binomial.

9. El pueblo de Punta Densidad ha sufrido, en promedio, un tornado por año en los últimos cien años, sin tendencia evidente

a) Dibuje un gráfico indicando el día en que Ud. espera el tornado en el lapso de un año.
b) ¿Cuál es la probabilidad de que ocurra el 4 de julio?
c) Asuma que los expertos consultados le indican que todos los últimos tornados han ocurrido entre julio y septiembre, siendo la fecha más probable el 15 de agosto. Grafique esta distribución.

10. Sobre la base de la información presentada en los estudios de caso de la entrega, que distribución seleccionaría para representar la concentración de radón en el interior de las viviendas, y la concentración de naftaleno en el aire?

<u>Apéndice</u>

Distribución Normal

$$f(x) = \frac{1}{\sigma . \sqrt{2\pi}} \; e^{-(x-\mu)^2 / 2\sigma^2}$$

Distribución Lognormal

$$f(x) = \frac{1}{x\sqrt{2\;\pi\sigma^2}} \; e^{-[en(x)-\mu]^2 / 2\sigma^2} \qquad \text{si } x > 0$$

Media $= e^{\mu + \sigma^2/2}$

Varianza $= e^{2\mu + \sigma^2 \left(e^{\sigma^2} - 1\right)}$

Distribución Triangular

$$f(x)\begin{cases} \dfrac{2(x-a)}{(b-a)(c-a)} & \text{si} \quad a \le x \le c \\ \dfrac{2(b-x)}{(b-a)(b-c)} & \text{si} \quad c \le x \le b \\ 0 \quad \text{en cualquier otro lugar} \end{cases}$$

Bibliografía

Burmaster, D. 1995. Invited paper in Human and Ecological Risk Assessment

Covelle, G. y Berger, R. 2002. Statistical Inference. Duxbury Press.

Kammen, D. y Hassenzahl, D. 2000. Should we Risk it?. Princeton University Press.

Evans, Q. y Olson, D. 1998. Introduction to Simulation and Risk Análisis. Prentice – Hall.

Morley, R.S. 1993. A model for the assessment of the animal disease risks associated with the importation of animals and animal products. Rev. Sci. Tech. OIE, 12 (4), 1055-1092.

Feller, W. 1980. Introducción a la teoría de probabilidades y sus aplicaciones. Ed. Limusa.

Freund, J. y Walpole, R. 1990. Estadística matemática con aplicaciones. Prentice-Hall Hispanoamericana, S.A..

Selvin, S. 1996. Statistical Análisis of Epidemiologic Data. Oxford University Press.

Vose, D. 2000. Risk Analysis: a Quantitative Guide. John Wiley and Sons.